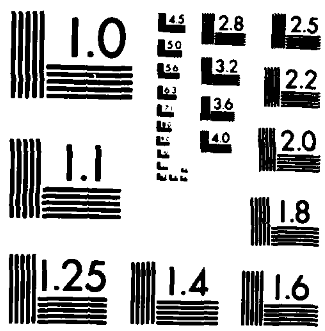
AD-A150 106   ASYMPTOTICS FOR CONFIGURAL ESTIMATORS(U) MASSACHUSETTS INST OF TECH CAMBRIDGE STATISTICS CENTER S MORGENTHALER SEP 84 TR-1 ARO-21779.1-MA   1/1

UNCLASSIFIED   DAAG29-84-K-0207   F/G 12/1   NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ARO 21779.1-MA

2

STATISTICS CENTER

Massachusetts Institute of Technology

111, Cambridge, Massachusetts 02139 (617) 253-872

AD-A150 106

ASYMPTOTICS FOR CONFIGURAL ESTIMATORS

BY

STEPHAN MORGENTHALER
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

TECHNICAL REPORT NO. ARO 1

SEPTEMBER 1984

PREPARED UNDER CONTRACT

DAAG29-84-K-0207

FOR THE U.S. ARMY RESEARCH OFFICE

DTIC FILE COPY

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

85 01 29 128

ASYMPTOTICS FOR CONFIGURAL ESTIMATORS

Stephan Morgenthaler

M.I.T.

Statistics Center

ABSTRACT

This paper examines the asymptotic properties of compromise estimators. By this we mean an estimation method which compromises between a finite number of sampling situations in a small sample optimal way. We develop the asymptotic theory of such estimators and show that under a specific choice of sampling situations the compromise estimator is asymptotically robust in Huber's sense.

ASYMPTOTICS FOR CONFIGURAL ESTIMATORS

Stephan Morgenthaler

M.I.T.

Statistics Center

1. Introduction

Configural polysampling denotes a method of estimation which is geared to small sample sizes and produces "robust" methods (see Pregibon and Tukey (1980)). There are important differences to the robustness philosophy as developed by Huber (see Huber (1964)). Since in small samples the distributions of estimators are quite intractable one has to rely on numerical methods in order to evaluate even relatively simple performance summaries like the mean-square-error. This holds true except in some simple cases -- like the Gaussian sampling model -- where a few expectations can be evaluated in closed form. In this connection it is important for the statistical community to realize that numerical methods are perfectly acceptable. They do, however, limit the number of sampling situations we can take into consideration. This in contrast to an asymptotic approach, where for simple models an infinity of sampling situations can be considered simultaneously (Huber (1964)).

Pitman (1938) for example solves the small sample problem for a single sampling situation in a location and scale setting. In this paper we will show what happens if Pitman's method is extended to two sampling situations. And we will address the question of the asymptotic performance of such estimators.

An asymptotic analysis is the simplest way to learn something about the behavior of an estimator in a variety of sampling situations. But it only gives a partial answer and we should not forget the more important approach based on performing "experiments" for small sample sizes. This paper, however, will restrict attention to asymptotic discussions.

In Section 2 we will introduce the idea of compromise estimators and discuss their optimality properties. Section 3 contains the corresponding asymptotic theory. As an example we define a compromise estimator which is asymptotically everywhere at least as good as Huber's minimax estimator.

2. Configural Estimators

2.1. Pitman's Estimator

Let $x_1, x_2, \ldots, x_n$ be n observations in an i.i.d. sampling situation from $F(x - \mu)$ where $1-F(x) = F(-x)$ for all x. We also assume that $F(x) \neq 0$ or 1 for any finite x and furthermore that F() has density f() with respect to Lebesgue measure.

We restrict attention to symmetric sampling situations in order to avoid the issue of what "parameter" we try to estimate. Symmetry of the underlying distribution allows us to define a target, namely μ = center of symmetry. Furthermore we will not get into any discussions if later on we allow for two -- or many -- different sampling situations. The center of symmetry is well defined for all symmetric shapes which means the estimation of μ is a well defined problem for a large class of sampling situations.

The solution Pitman gives is

$$T_P(x_1,\ldots,x_n) = -\frac{\int_{-\infty}^{\infty} r \prod_{i=1}^{n} f(x_i + r)\, dr}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} f(x_i + r)} \qquad (2.1)$$

(see Pitman (1938)). This estimator has the smallest mean-square-error among all location equivariant estimators. Location equivariance is a very reasonable restriction on a location estimator T(), it means that

$$T(x_1 + r,\ldots,x_n + r) = T(x_1,\ldots,x_n) + r,\ r \in R, \qquad (2.2)$$

i.e. the estimator changes in the same way as the sample.

remarks: (1) The most revealing way of deriving (2.1) is through the concept of "configurations". By this notion we mean the pattern of the points in our (ordered) sample and it is easily seen that this is an ancillary statistic. The Pitman estimator then is chosen such that conditioned on the configuration the estimate is unbiased. Since the conditional variance can not be affected by the choice of the estimate (under equivariance), this has to produce the smallest mean-square-error.

(2) The conditions on f() such that (2.1) exists are discussed in Pitman (1938).

Formula (2.1) produces an estimator T_F of the center of symmetry μ no matter what the underlying sampling situation. It therefore need not be so that the x_i's are sampled from $F(x-\mu)$.

Let us therefore introduce $G(x-\mu)$ -- again $G(x) = 1 - G(-x)$ for all x's -- as the sampling situation for $x_1, \ldots, x_n$. This is a new way of looking at the Pitman estimator T_F and it of course immediately lets us see the optimality property in a new light. If e.g. $F = \Phi$ and G = Cauchy we are looking at the behavior of the arithmetic mean under Cauchy sampling. If we are open minded about the assumptions we base our inference on, we have to admit that in small samples we can not with any reasonable precision know what the underlying sampling situation is nor should we attempt to make inferences about it. Huber (1964) formalizes the idea of a robust method as a procedure which "behaves well" in the neighborhood of a parametric model. Huber therefore would allow G() to be chosen somewhere near F() and he modifies T_F in such a way that the behavior of the new estimate is acceptable for all allowed G()'s. This leads us away from considering estimates like T_F which are optimized at a single "point". Since -- in small samples -- we will never be able to tell at which "point" we are, it ought to be obvious that single-point-optimization is a bad strategy.

2.2. Compromise Estimators

Let us now consider the case where $x_1, \ldots, x_n$ is a sample from either $F_1(x-\mu)$ or $F_2(x-\mu)$, where F_1 and F_2 satisfy all the constraints of F (see beginning of Section 2.1). We are now interested in location equivariant estimators which optimize at two "points", namely F_1 and F_2, simultaneously. This is obviously impossible. However, decision theory teaches us that estimates of the form

$$T_{F_1, F_2, \alpha}(x_1, \ldots, x_n) =$$

$$= \frac{\int r \left\{\alpha \prod_{i=1}^{n} f_1(x_i + r) + (1-\alpha) \prod_{i=1}^{n} f_2(x_i + r)\right\} dr}{\int \left\{\alpha \prod_{i=1}^{n} f_1(x_i + r) + (1-\alpha) \prod_{i=1}^{n} f_2(x_i + r)\right\} dr} \tag{2.3}$$

($0 < \alpha < 1$) are bi-optimal in the sense that they cannot be improved in both sampling situations F_1 and F_2 simultaneously (see Ferguson (1968)).

remarks: (1) We can also write

$$T_{F_1, F_2, \alpha}(x_1, \ldots, x_n) =$$

$$\alpha\, w_{F_1}(x_1, \ldots, x_n)\, T_{F_1}(x_1, \ldots, x_n) +$$

$$(1-\alpha)\, w_{F_2}(x_1, \ldots, x_n)\, T_{F_2}(x_1, \ldots, x_n) \tag{2.4}$$

where

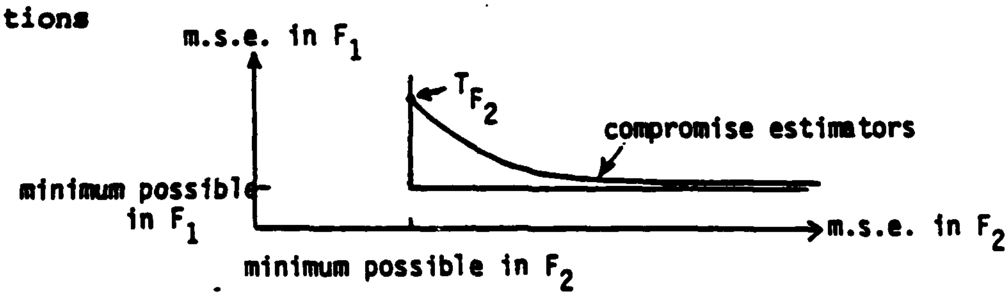
$$w_{F_k}(x_1, \ldots, x_n) = \frac{\int \prod_{i=1}^{n} f_k(x_i + r)\, dr}{\int \{\alpha \prod_{i=1}^{n} f_1(x_i + r) + (1-\alpha) \prod_{i=1}^{n} f_2(x_i + r)\}\, dr}$$

$(k = 1, 2)$ and $T_{F_k}()$ is defined in (2.1). We therefore can interpret the family of bi-optimal estimators as a weighted mean of the single-situation optimal estimators. Note, however, that the weights are "adaptive", they depend on the sample values. Of course any equivariant estimator can be represented as a weighted mean of the single-situation optimal estimators. What matters here is the simplicity and form of the weights together with their small sample optimality property.

(2) It is clear from (2.3) that $T_{F_1, F_2, 0} = T_{F_2}$ and $T_{F_1, F_2, 1} = T_{F_1}$.

(3) The picture which helps us most in understanding the compromise estimators is shown in Figure 2.1.

Figure 2.1: plot of the mean-sqare-errors in the two situations



Note that since we only consider location equivariant estimators the risk in any given situation does not depend on

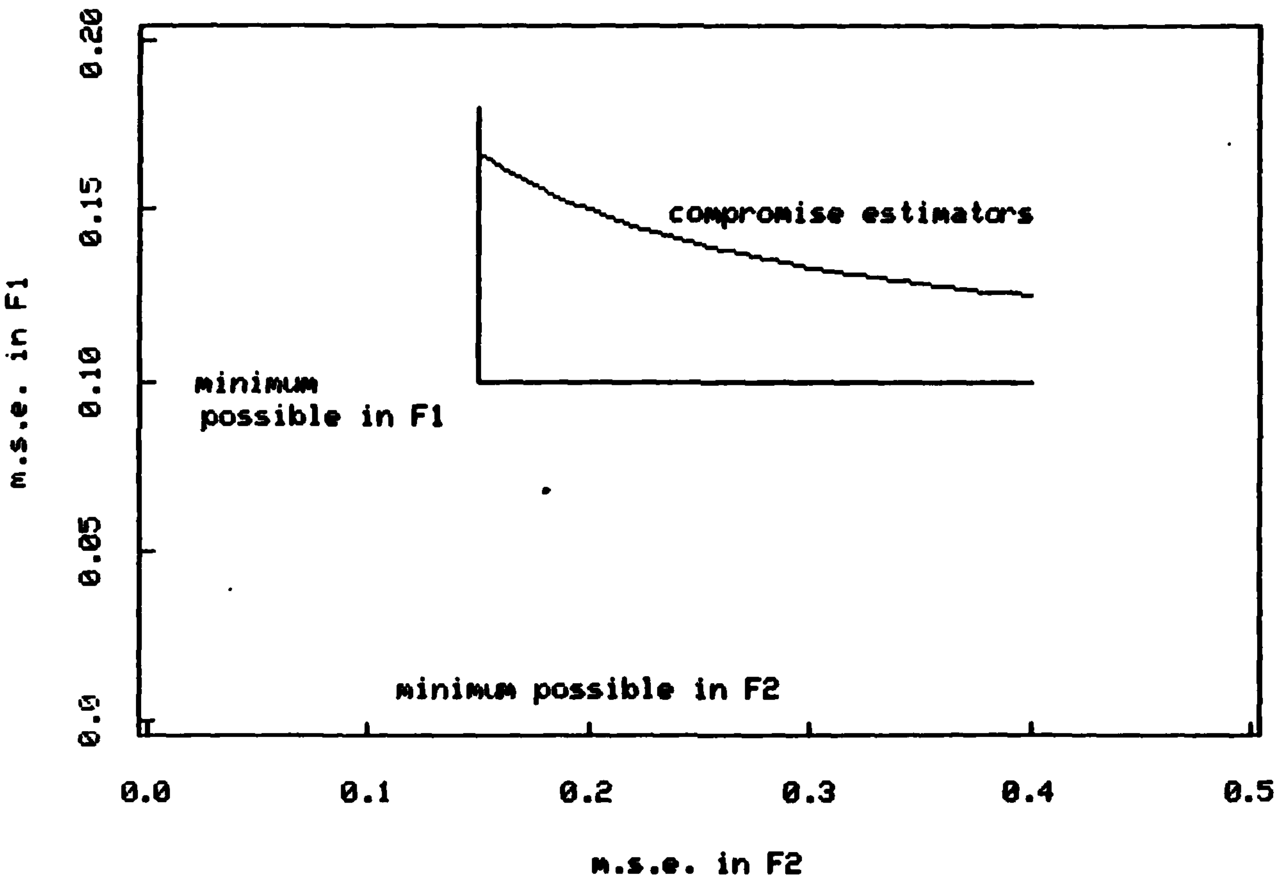


Figure 2.1: Plot of the mean-square-error in the two situations

the parameter value μ (see Ferguson (1968)). The bi-optimal or compromise estimators are the ones which lie on the convex boundary curve.

(4) A Bayesian interpretation of the estimator (2.3) is possible. In that framework $(\alpha, 1 - \alpha)$ is a prior distribution on the set of underlying sampling shapes.

(5) In order to implement (2.3) in an actual application, the formula (2.4) has some interesting interpretations. Pregibon and Tukey (1980) derive the formulas from the point of view of sampling. This leads to the consideration of different weights w_{F_1} and w_{F_2}.

The choice of the two compromising distributions F_1 and F_2 is of importance in actual applications of the technique. In many applications it is traditional to consider $F_1 = \Phi$, the Gaussian shape. The choice of F_2 is somewhat related to the choice of the contamination parameter ϵ in Huber's model. F_2 will influence two aspects (see (2.4)):

(i) the "relative weights" w_{F_1} and w_{F_2}

(ii) the "other" optimal estimator T_{F_2}.

These two aspects have an interpretation in the theory of M-estimators. The first is connected with the choice of tuning constants, like k in Huber's ψ_k-function ($\psi_k(x) = \max(-k,\min(k,x))$), and the second with the shape of the ψ-function. From small sample studies we know for example that a redescending ψ-function is advantageous -- it costs little and buys a lot. This is reflected in the theory of compromise estimators by the choice of F_2 and by how far F_2 is away from Φ.

3. The Asymptotic Behavior of Compromise Estimators

In this section we are going to explore what happens to compromise estimators (see (2.3) or (2.4)) if we sample from a distribution G() and let the sample size n grow. We will see that the weights w_{F_1} and w_{F_2} usually tend to (0,1) or (1,0), respectively. A compromise estimator for large sample sizes therefore will be close to either the optimal estimate under F_1 or the optimal estimate under F_2. This is a reasonable behavior since the "information" about the sampling situation G() grows as the sample size gets large. The distinction between F_1 and F_2 is therefore more and more estimable. In a few words then, we can say that compromise estimators exhibit an adaptive behavior with the relative weights w_{F_1} and w_{F_2} (see (2.4)) gauging the adaptation.

3.1. The Asymptotic Behavior of the Relative Weights

Suppose $x_1, \ldots, x_n$ is a sample of size n from $G(x-\mu)$. We assume that G() is symmetric around 0. The relative weights are defined as

$$w_{F_k}(x_1, \ldots, x_n) = \frac{\int \prod_{i=1}^{n} f_k(x_i + r)\, dr}{\int \{\alpha \prod_{i=1}^{n} f_1(x_i + r) + (1-\alpha) \prod_{i=1}^{n} f_2(x_i + r)\}\, dr} \qquad (3.1)$$

($k = 1$ or 2), where the notation is the same as in (2.4). The following Lemma treats an "overly nice" case.

Lemma 3.1

Let us assume that both $-\log f_1(x)$ and $-\log f_2(x)$ are convex. And let us furthermore assume that $G()$ is such that the functions

$$A^1(r) = \int \log f_1\,(x + r)\;dG(x)$$

and

$$A^2(r) = \int \log f_2\,(x + r)\;dG(x)$$

exist for all r and achieve a unique maximum at $r=0$. If

$$\int \log f_1(x)\;dG(x) > \int \log f_2(x)\;dG(x) \qquad (3.2)$$

it follows that

$$\frac{w_{F_2}(x_1,\ldots,x_n)}{w_{F_1}(x_1,\ldots,x_n)} \longrightarrow 0 \text{ a.s.}$$

proof: Let $X_1, X_2, \ldots$ denote a sequence of iid random variables with common distribution $G(x)$. From (3.1) we have

$$\frac{w_{F_2}(X_1,\ldots,X_n)}{w_{F_1}(X_1,\ldots,X_n)} = \frac{\int \prod_{i=1}^{n} f_2(X_i + r)\;dr}{\int \prod_{i=1}^{n} f_1(X_i + r)\;dr}.$$

Now

$$I(X_1,\ldots,X_n) = \int \prod_{i=1}^{n} f(X_i + r)\;dr$$

$$= \int \exp\Big(n\Big(\frac{1}{n}\sum_{i=1}^{n} \log f(X_i + r)\Big)\Big)\;dr$$

$$= \int \exp(n\, A_n(r))\, dr,$$

where

$$A_n(r) = \frac{1}{n} \sum_{i=1}^{n} \log(f(X_i + r))$$

and f stands for either f_1 or f_2.

Since we are interested in the large sample behavior of $I(X_1,\ldots,X_n)$ we can use an asymptotic expansion argument to approximate I().

We know that for large n

$$I(X_1,\ldots,X_n) \sim \int \exp(n\{A_n(R_0^n) - \frac{1}{2} A''_n(R_0^n)\ (r - R_0^n)^2\})\ dr$$

$$\sim \exp(nA_n(R_0^n))\ (2\frac{\pi}{n})^{\frac{1}{2}} \frac{1}{(A''_n(R_0^n))^{\frac{1}{2}}} \qquad (3.3)$$

where R_0^n denotes the point where the (random) function $A_n()$ takes its maximal value. Such a single maximum exists because of our convexity assumptions.

The theory of asymptotic expansions is treated for example in Chapter 6 of Dingle (1973).

If we blend the probability structure which underlies the sequence $X_1, X_2,\ldots$ (due to iid sampling from G()) with the asymptotic approximation (3.3), we can say something about the behavior of the right hand side in (3.3). Because of the strong law of large numbers and our assumptions we have

$$A_n(r) \dashrightarrow A(r) \quad \text{a.s.} \quad \text{for all } r \qquad (3.4)$$

and from this we can conclude that

$$R_0^n \dashrightarrow 0 \quad a.s.$$

so that

$$\frac{1}{n} \log I(X_1,\ldots,X_n) \dashrightarrow A(0) = \int \log f(x)\, dG(x) \quad a.s.$$

where f() denotes either $f_1()$ or $f_2()$ (see (3.3)).

We therefore conclude from

$$\int \log f_1(x)\, dG(x) > \int \log f_2(x)\, dG(x)$$

that

$$\frac{1}{n} \log I_1(X_1,\ldots,X_n) - \frac{1}{n} \log I_2(X_1,\ldots,X_n)$$

$$= (-\frac{1}{n}) \log \left(\frac{w_{F_2}(X_1,\ldots,X_n)}{w_{F_1}(X_1,\ldots,X_n)}\right) \dashrightarrow \text{constant} > 0 \quad a.s.$$

where $I_1(X_1,\ldots,X_n)$, $I_2(X_1,\ldots,X_n)$ refer to $f() = f_1()$, $f() = f_2()$, respectively.

From this last statement the theorem follows immediately.

Lemma 3.1 is not good enough for our purposes since the assumption about $A^1(r)$ is not always satisfied. If for example $f_1 = \phi$ and G() has an infinite second moment then $A^1(r) = -\infty$ for all r. But we would of course still expect that Lemma 3.1 holds, in fact we would hope for `very fast convergence of the ratio of relative weights.

<u>Lemma</u> <u>3</u>.<u>2</u>

Let us assume that both $-\log f_1(x)$ and $-\log f_2(x)$ are convex. And let us furthermore assume that G() is such that the function

$$A^2(r) = \int \log f_2 \; (x + r) \; dG(x)$$

exist for all r and achieves a unique maximum at r=0. If $\int \log f_1 \; (x) \; dG(x) = -\infty$ then it follows that

$$\frac{w_{F_2}(x_1,\ldots,x_n)}{w_{F_1}(x_1,\ldots,x_n)} \longrightarrow 0 \text{ a.s.}$$

__proof__: Use the same argument as in the proof of Lemma 3.1 and note that

$$\frac{1}{n} \log I_1(X_1,\ldots,X_n) < \text{any constant} \quad \text{a.s.}$$

__remarks__: (1) The asymptotic expansion (3.3) shows how closely the maximum likelihood estimator is connected to the Pitman estimator. Note that $A''(0)$ is equal to the Fisher information if $G() = F()$, i.e. the sampling situation and the modelling situation agree. We will see below that the maximum likelihood estimator T_{F_1} is indeed asymptotically identical to the Pitman estimator.

(2) It is reasonable to believe that Lemma 3.1 holds in greater generality. The convexity conditions on the log densities are probably not needed.

__Corollary 3.1__ Under the assumptions of the Lemma 3.1 or Lemma 3.2 and if

$$\int \log f_1(x) \; dG(x) > \int \log f_2(x) \; dG(x)$$

then the compromise estimator $T_{F_1,\, F_2,\, \alpha}$ ($\alpha \neq 0$) is asymptotically equivalent to the Pitman estimator T_{F_1}.

__proof__: Apply the Lemmas to Formula (2.4) remarks: (1) Corollary 3.1 states that as the sample size increases the compromise estimator will pick either one of the two single-situation-optimal estimates depending on (3.2).

We therefore expect that

$$\int \log f_1(x)\, dG(x) - \int \log f_1(x)\, dG(x)$$

$$= \int \log \left(\frac{f_1(x)}{f_2(x)}\right) dG(x) \qquad (3.5)$$

is a quantity which decides whether the sampling situation G is "closer" to the modelling situation F_1 or the modelling situation F_2.

The quantity (3.5) is closely related to the Kullback-Leibler mean information for discrimination (Kullback and Leibler (1951)). Their formula is

$$I(1:2) = \int \log\left(\frac{f_1(x)}{f_2(x)}\right) f_1(x)\, dx,$$

where I(1:2) is the mean information for discrimination per observation from sampling situation F_1.

(2) The asymptotic behavior of the compromise estimators (2.3) does not depend on α (unless (3.5) = 0).

(3) More results about Pitman estimators can be found in Johns (1979) and Klaassen(1981). George Easton has proved the results given in Section 3.1 for the more general case of unknown scale (Easton (1984)).

3.2. Asymptotics of the Pitman Estimator

In order to get asymptotic efficiencies for the compromise estimators we need to know more about the asymptotic behavior of the Pitman estimators T_{F_1} and T_{F_2}. Port and Stone (1974) provide the information in the case where the sampling situation and the modelling situation are identical. In our more general setup we can argue the following way:

$$T_F(x_1,\ldots,x_n) = -\frac{\int r\ \exp(nA_n(r))\ dr}{\int \exp(n\ A_n(r))\ dr},$$

where

$$A_n(r) = \frac{1}{n}\sum_{i=1}^{n} \log f(x_i + r) \quad (f() = \frac{d}{dx} F()).$$

If we expand the numerator asymptotically we get

$$T_F(x_1,\ldots,x_n) \simeq$$

$$-\frac{\exp(n\ A_n(r_0^n)) \int r\ \exp(-\frac{n}{2}A''_n(r_0^n)\ (r - r_0^n)^2)}{\int \exp(n\ A_n(r))\ dr}$$

$$\simeq - r_0^n$$

where r_0^n maximizes $A_n(r)$. We therefore showed that asymptotically the Pitman estimator and the maximum likelihood estimator $(- r_0^n)$ agree. This agreement is good enough to conclude that the asymptotic distributions are the same. Huber (1967) then provides the necessary results.

3.3. Huber's Contamination Model: An Example

To illustrate the use of the theory we developed let us look at the compromise estimators based on the two modelling densities

$$f_1(x) = \phi(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \exp(-\tfrac{1}{2}x^2)$$

$$f_2(x) = (1-\alpha)\ \phi(x) \qquad \text{if } |x| < k$$

$$= \frac{(1-\alpha)}{(2\pi)^{\frac{1}{2}}} \exp(\tfrac{k^2}{2} - k|x|) \quad \text{otherwise}$$

(where k is such that $\frac{2\phi(k)}{k} - 2\Phi(-k) = \frac{\alpha}{1-\alpha}$).
The alternative density is of course the least favorable choice in the class of distributions $\{(1-\alpha)\Phi() + \alpha H() : H() \text{symmetric}\}$ (see Huber (1964)).

The asymptotic variance of an estimator compromising between these two symmetric situations (see (2.3)) will be equal to either of the asymptotic variances of the Pitman estimators

T_{F_1} = arithmetic mean

or

T_{F_2} = Pitman estimator for the least favorable distribution.

If we sample from distribution G() we have for these asymptotic variances ($\mu_G = \int x \, dG(x)$)

$$\text{as. var}_G(T_{F_1}) = \int (x - \mu_G)^2 \, dG(x)$$

$$\text{as. var}_G(T_{F_2}) = \frac{\int (\psi_k(x - \mu_G))^2 \, dG(x)}{(\int \psi'_k(x - \mu_G) \, dG(x))^2}$$

where $\psi_k(x) = -f'_2(x)/f_2(x) = \max(-k, \min(k,x))$.

In his 1964 paper Huber shows that the M-estimator based on $\psi_k()$ is asymptotically minimax for sampling situations chosen from the contamination class. Since T_{F_2} has the same asymptotic behavior as this M-estimator, the same claim can be made for T_{F_2}. Note, however, that for finite sample sizes T_{F_2} will be superior. The following Proposition explains the asymptotic behavior of the compromise estimators (see (2.3)).

<u>Proposition</u> <u>3</u>.<u>1</u> Let $G(x) = (1-\alpha)\Phi(x) + \alpha H(x)$ where $H(x) + H(-x) = 1$ for all x's and H() puts all its mass outside the interval $[-k,k]$, but is otherwise arbitrary. Furthermore assume that $0 \leq \alpha \leq 0.5$. Then

as. var_G(compromise estimator) $\leq$ as. var_G(Huber's minimax estimator)

<u>proof</u>: From Lemma 3.1 and Lemma 3.2 we know that

$$\int \log f_1(x)\, dG(x) - \int \log f_2(x)\, dG(x)$$

$$= \int \{\log \frac{1}{(2\pi)^{\frac{1}{2}}} - \tfrac{1}{2}x^2\}\, dG(x) - \int \log \left(\frac{1-\alpha}{(2\pi)^{\frac{1}{2}}}\right) dG(x)$$

$$- \int_{-k}^{k} \tfrac{1}{2}x^2\, dG(x) - 2\int_{k}^{\infty} \left(\frac{k^2}{2} - k|x|\right) dG(x)$$

$$= -\log(1-\alpha) + 2\int_{k}^{\infty} \{k|x| - \frac{k^2}{2} - \frac{x^2}{2}\}\, dG(x) \qquad (3.6)$$

is the quantity which decides about the asymptotic variance of the compromise estimator. Note that we made use of the symmetry of the sampling distribution G in the derivation of (3.6). If (3.6) is positive the compromise estimators will behave asymptotically like the arithmetic mean, otherwise like the Huber-estimator. All that remains to be considered therefore is the case where (^.6) is positive (or zero) because in the other case the assertion of the Proposition is trivial.

First note that (3.6) can only be positive if G has finite variance. Using our assumptions about $G = (1-\alpha)\Phi + \alpha H$ stated in the Proposition, (3.6) can be written as

$$-\log(1-\alpha) + 2(1-\alpha)\int_k^{\infty}\left(k|x| - \frac{k^2}{2} - \frac{x^2}{2}\right)\phi(x)\,dx +$$

$$\alpha\int\left(k|x| - \frac{k^2}{2} - \frac{x^2}{2}\right)dH(x)$$

$$= -\log(1-\alpha) - (1-\alpha)\int_k^{\infty}(x-k)^2\,d\Phi(x) - \frac{\alpha}{2}\int(x-k)^2\,dH(x)$$

$$= -\log(1-\alpha) - (1-\alpha)\,\{k\phi(k) + \Phi(-k)(1+k^2) - \frac{\alpha}{2}k^2 - \frac{\alpha}{2}\sigma_H^2 \tag{3.7}$$

where $\sigma_H^2 = \int x^2\,dH(x)$ is the variance of the contaminating distribution.

A comparison of the asymptotic variances of the sample mean and Huber's estimator is not hard. We have

$$\text{as. var}_G(\text{sample mean}) = (1-\alpha) + \alpha\,\sigma_H^2 \tag{3.8}$$

$$\text{as. var}_G(\text{Huber-estimator}) = \frac{\int(\psi_k(x))^2\,dG(x)}{(\int\psi'_k(x)\,dG(x))^2}$$

$$= \frac{\int_{-k}^{k} x^2 \, dG(x) + 2 \int_{k}^{\infty} k^2 \, dG(x)}{\left(\int \psi_k(x) \; dG(x)\right)^2}$$

$$= \frac{(1-\alpha) \int_{-k}^{k} x^2 \, d\Phi(x) + k^2\alpha + 2k^2(1-\alpha)\Phi(-k)}{(1-\alpha)^2 \; (\Phi(k) - \Phi(-k))^2} \qquad (3.9)$$

In this last formula we have again used all our knowledge about the sampling situation G().

What remains to be shown is

non-negativeness in (3.7) --> 3.8 $\leq$ (3.9).

But (3.7) $\geq$ 0 -->

$$\alpha\sigma_H^2 \leq -2 \log(1-\alpha) + 2(1-\alpha)k\phi(k) - 2(1-\alpha)\Phi(-k)(1+k^2) - \alpha k^2$$

and therefore we have (3.8) = $(1-\alpha) + \alpha \, \sigma_H^2 \leq$

$$(1-\alpha) + \log\left(\frac{1}{1-\alpha}\right)^2 + 2(1-\alpha)k\phi(k) - 2(1-\alpha)\Phi(-k)(1+k^2) - \alpha k^2$$

Using the equation linking α and k

$$\frac{2\phi(k)}{k} - 2\Phi(-k) = \frac{\alpha}{(1-\alpha)}$$

we can simplify and get

$$(3.8) \leq \log\left(\frac{1}{1-\alpha}\right)^2 + (1-\alpha)k^2\frac{\alpha}{1-\alpha} + (1-\alpha) - 2(1-\alpha)\Phi(-k) - \alpha k^2$$

$$(3.8) \leq \log\left(\frac{1}{1-\alpha}\right)^2 + (1-\alpha)(1-2\Phi(-k))$$

along the same line of thoughts we can simplify (3.9) to get

$$(3.9) = \frac{1}{(1-\alpha)(1 - 2\Phi(-k))}.$$

Putting all these results together we finally have

$$(3.9) \geq 1 + (1-\alpha)\frac{2\phi(k)}{k}$$

$$\geq (1-\alpha)(1-2\bar{\Phi}(-k)) + (1-\alpha)\frac{4\phi(k)}{k}$$

$$\geq (1-\alpha)(1-2\bar{\Phi}(-k)) + \log(\frac{1}{1-\alpha})^2 \geq (3.8)$$

if only we show that

$$(1-\alpha)\frac{4\phi(k)}{k} \geq \log(\frac{1}{1-\alpha})^2 \qquad (3.10)$$

holds. This last inequality is only true for α small enough, e.g. $\alpha \leq 0.5$. For such α values we have

$$\log(\frac{1}{1-\alpha})^2 \leq 3\alpha \quad (0 \leq \alpha \leq 0.5)$$

and (3.10) is therefore proved if we show that

$$(1-\alpha)\frac{4\phi(k)}{k} \geq 3\alpha$$

$$\Longleftrightarrow \frac{4\phi}{3}\frac{(k)}{k} \geq \frac{\alpha}{1-\alpha} = 2\frac{\phi(k)}{k} - 2\bar{\Phi}(-k)$$

$$\Longleftrightarrow 2k\bar{\Phi}(-k) \geq \frac{2}{3}\phi(k)$$

$$\Longleftrightarrow 3k\bar{\Phi}(-k) \geq \phi(k) \quad \text{for all } .436 \leq k < \infty. \qquad (3.11)$$

Note that the range of α values from 0 to 0.5 translates into a range of values for k.

This last inequality (3.11) which is equivalent to (3.10) does indeed hold and is left for the reader to check.

Proposition 3.1 is now proved for all the cases where (3.6) is strictly positive. Some care is needed if (3.6) is zero. Then the compromise estimator is asymptotically a convex linear combination of T_{F_1} and T_{F_2}, but since the asymptotic variance of T_{F_1} is lower than the asymptotic variance of T_{F_2} the compromise estimator will have an asymptotic variance below the asymptotic variance of T_{F_2}.

remarks: (1) We have identified a class of sampling situations G, namely those where (3.6) is positive, for which the mean is more efficient estimator than Huber's minimax estimator. It would be of interest to show how big this class is and also to check whether it contains all sampling situations for which the sample mean is asymptotically better than Huber's minimax estimator.

4. Discussion

This paper deals with estimators which compromise between different "shapes". This idea, as we have seen, produces "robust" estimators. If we compromise between the Gaussian and Huber's least favorable distribution we have a family of estimators (for different values of α) which dominate Huber's minimax M-estimator asymptotically.

Several points need to be clarified, however. The idea of compromising is different from the usual asymptotic robustness theory as developed by Huber (see Huber (1964) and Huber (1981)). There, the compromising takes place in a neighborhood of the "central" model, whereas in our approach the different shapes need not be close together. A neighborhood model is in fact only a first step towards robust/resistant techniques for small sample sizes. For samples of size 5 we would advise to compromise between the Gaussian and something like the slash (= distribution of a ratio of a Gaussian over an independent uniform) rather than using the only moderately tailed least favorable distribution.

The intention of this paper is not to show that we should use a compromise between the Gaussian and the least favorable distribution, but rather to let people know of the merits of compromise estimators in a language which many statisticians are used to, namely asymptotics.

Results found through small sample experiments are of greater importance. It is clear for example that the situations (or shapes) we compromise ought to change with the sample size. The amount of "information" in the sample grows with the sample size. Not only are we able to estimate "parameters" with less variability, we also gain insight into the underlying shape. Compromise estimators use this knowledge in an optimal way and with our choice of the shapes we can fine-tune the procedure. Important choices have to be made in that respect and more (probably experimental) research for small sample sizes is needed. Subject matter knowledge might prove useful in this connection.

The extension of Pitman's ideas to more than one shape provides us with a tool to find meaningful small sample methods of the robust/resistant kind. In order to make the asymptotics simple we did not deal with the scale parameter. In actual applications the inclusion of this additional parameter is, however, no problem (see Bell and Morgenthaler (1981) for an example).

References:

Bell Krystinik, K. and Morgenthaler, S. (1981): "Comparison of the Bioptimal Curve with Curves for Two Robust Estimates". Technical Report No. 195, Series 2, Department of Statistics, Princeton University

Dingle, R.B. (1973): Asymptotic Expansions: Their Derivation and Interpretation. Academic Press, London and New York

Easton, G. (1984): Personal Communication

Ferguson, T.S. (1967): Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York

Huber, P.J. (1964): "Robust Estimation of a Location Parameter". Ann. Math. Statist., 35, 73-101

Huber, P.J. (1967): "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions". in: Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California Press, Berkeley

Huber, P.J. (1981): Robust Statistics. Wiley, New York

Johns, M.V. (1979): "Robust Pitman-like Estimators". in: Robustness in Statistics edited by R.L. Launer and G.N. Wilkinson, Academic Press, New York

Klaassen, C.A.J. (1981): Statistical Performance of Location Estimators. Mathematical Centre Tracts, No. 133, Amsterdam

Kullback, S. and Leibler, R.A. (1951): "On Information and Sufficiency". Ann. Math. Statist., 22, 79-86

Pitman, E.J.G. (1938): "The Estimation of the Location and Scale Parameters of a Continuous Population of Any Given Form". Biometrika, 30, 391-421

Port, S.C. and Stone, C.J. (1974): "Fisher Information and the Pitman Estimator of a Location Parameter". Ann. Statist., 2, 225-247

Pregibon, D. and Tukey, J.W. (1981): "Assessing the Behavior of Robust Estimates of Location in Small Samples: Introduction to Configural Polysampling". Technical Report No. 185, Series 2, Department of Statistics, Princeton University

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 21779.1-MA	2. GOVT ACCESSION NO. AD-A150 106	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Asymptotics For Configural Estimators		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stephan Morgenthaler		8. CONTRACT OR GRANT NUMBER(s) DAAG29-84-K-0207
9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistics Center Massachusetts Institute of Technology Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N/A
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE September 1984
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE;DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

N/A

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robust estimation, conditional inference, equivariance, asymptotics,

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper examines the asymptotic properties of compromise estimators. By this we mean an estimation method which compromises between a finite number of sampling situations in a small sample optimal way. We develop the asymptotic theory of such estimators and show that under a specific choice of sampling situations the compromise estimator is asymptotically robust in Huber's sense.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

3-85

DTIC